Adaptive Online Documents

David Salesin
Chuck Jacobs
Wilmot Li

The Future of PC

Rich Media Experiences

Three factors

- The growth of the imaging and graphics capabilities on the PC
- The explosion of consumer end digital image and video acquisition
- The increase in connectivity

PC Visual Media Experience Today

- Passive viewing of video and imagery
- Browsing and editing based on a linear time-line
- Indexing and searching not sensitive to content
- Annotating and story telling primitive
- Each media type lives in its own “box”

Media, Interaction, and Interactivity

Our goal

Revolutionize the consumer PC media experience by enabling and enhancing:

Interactivity

- Content that can be touched and is reactive

Social Interaction

- Sharing experiences, reliving memories through media

A Three Part Agenda

- Media content analysis
  - Novel representations, intelligent processing
- Enhanced viewing and sharing pof today's media
  - Documents, presentations, annotations, and video
- New types of media experiences
  - Interactive content and integrated media

Direct Impact on MS Products

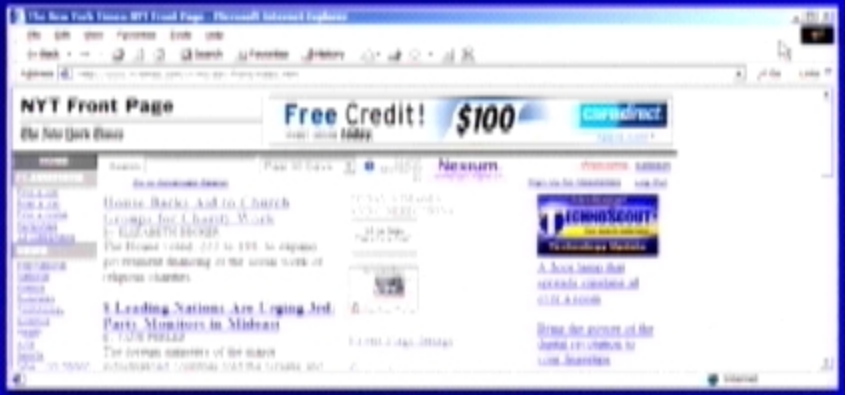
- Layout and reading experiences
  - eDocs, TabletPc
- Annotations and reflow
  - Office, Avalon/eDocs, TabletPC/RichInk
- Video
  - Shell MSX, Windows Media Tools
- IBR
  - Avalon

State of the art

State of the art

State of the art

1. Impoverished layout.
2. Very limited ability to adapt.

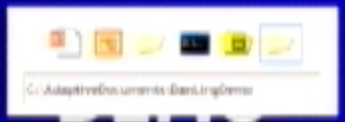


In Contrast: The Real Thing

"Adaptive documents"

- Two key ideas:

1. Multiple representations for all content (sections, words, images, links, buttons, etc.)
2. Content chosen and formatted dynamically to fit the viewing situation (audience, device size, bandwidth, etc.)



INSIDE THE WAR ROOM

By James Carney and John F. Dickerson

D

INSIDE THE WAR ROOM

By James Carney and John F. Dickerson

D

INSIDE THE WAR ROOM

By James Carney and John F. Dickerson

D

INSIDE THE WAR ROOM

By James Carney and John F. Dickerson

D

INSIDE THE WAR ROOM

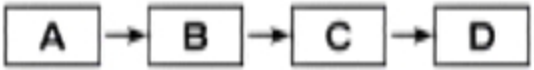
Adaptable Animated Presentations

Doug Zongker, Tomer Moscovich, Karin Scholz, John Hughes, David Salesin

Microsoft Research, Univ. of Washington, Brown Univ.

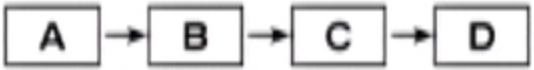
static organization

talk.ppt

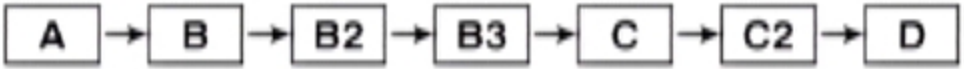


static organization

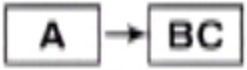
talk.ppt



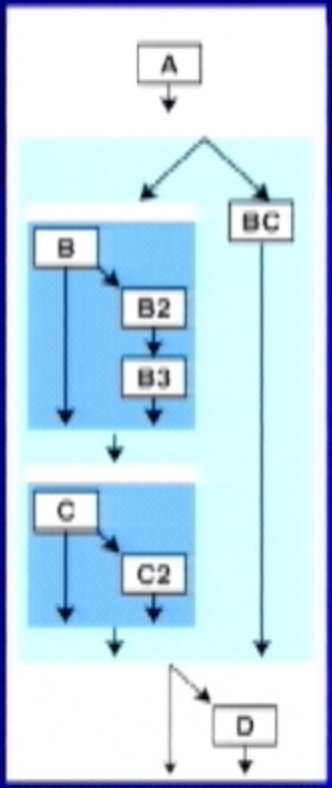
talk-extended.ppt



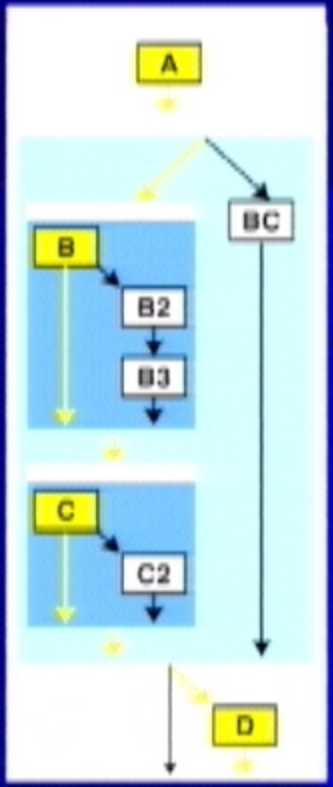
talk-summary.ppt



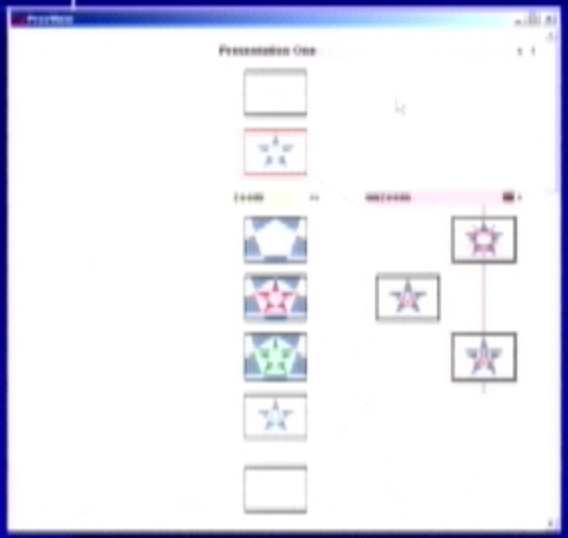
nonlinear structure



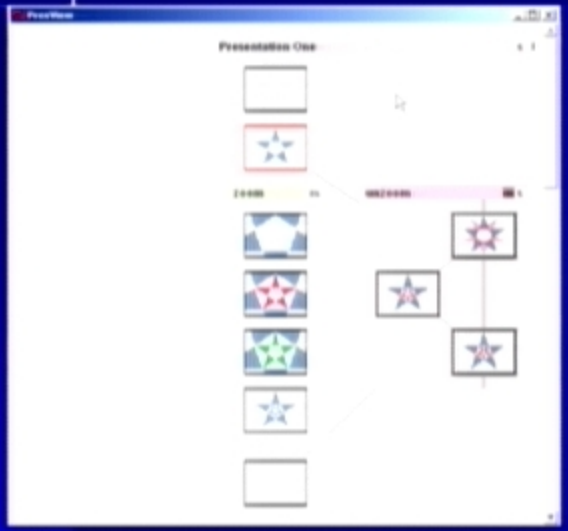
nonlinear structure

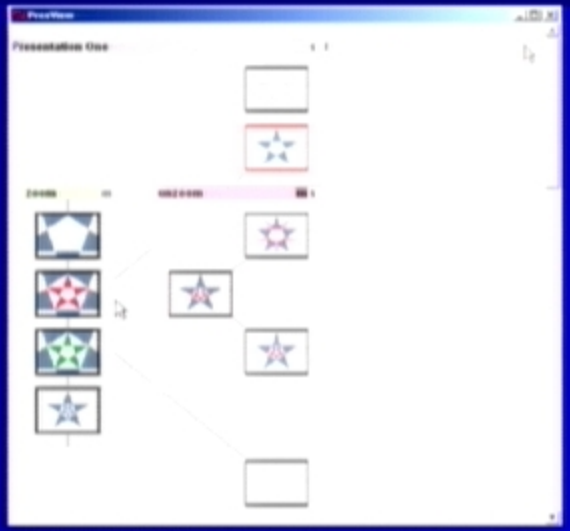


navigation view



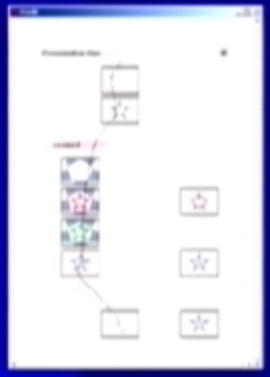
navigation view









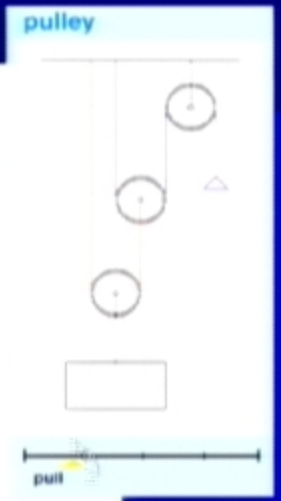


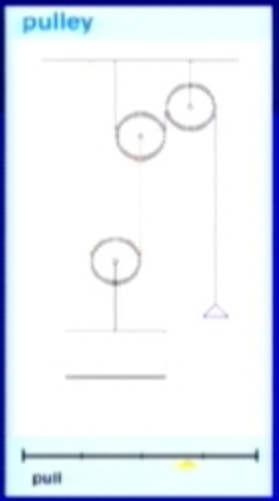
animated content

animated content

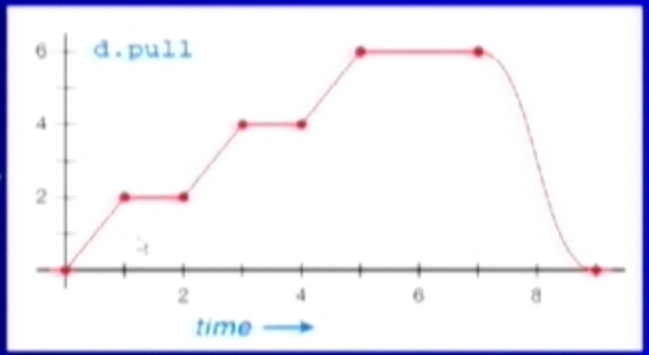
- other
- than
- flying
- bullet
- points

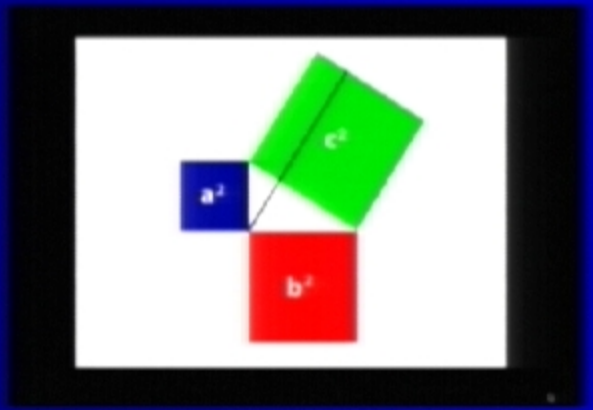
```
def pulley( pull=(0,6) ):
    b_lift = pull/2.0
    c_lift = pull/4.0

    clear( 1, 1, 1 )

    thickness( 3 )
    color( 0 )
    line( (0,15), (8,15) )

    push()
    translate( 0, c_lift )
    stroke( weight )
    pop()

    color( 0, 0, 0.7 )
    line( (3,15), (3,9+b_lift) )
    line( (5,9+b_lift), (5,13) )
```

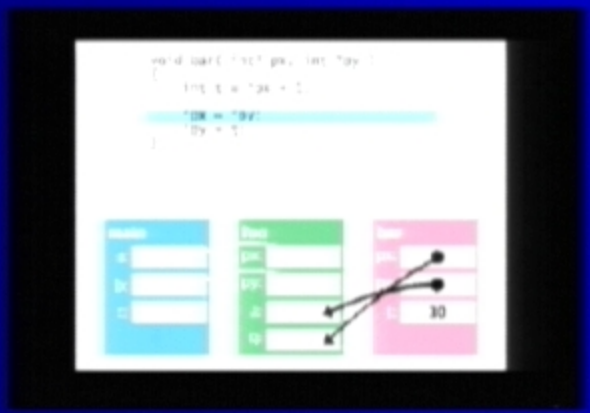




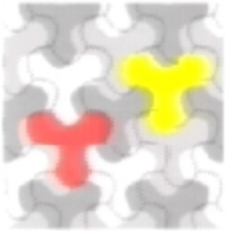
```
d = Diagram( ..., draw = pulley )

start_animation( [d], 'slide_1' )

# set initial position of weight
set( d.pull, 0.0 )

# increase pull in small steps
linear( 1.0, d.pull, 2.0 )
wait( 1.0 )
linear( 1.0, d.pull, 4.0 )
wait( 1.0 )
linear( 1.0, d.pull, 6.0 )
wait( 2.0 )

# smoothly lower weight back down
smooth( 2.0, d.pull, 0.0 )
```

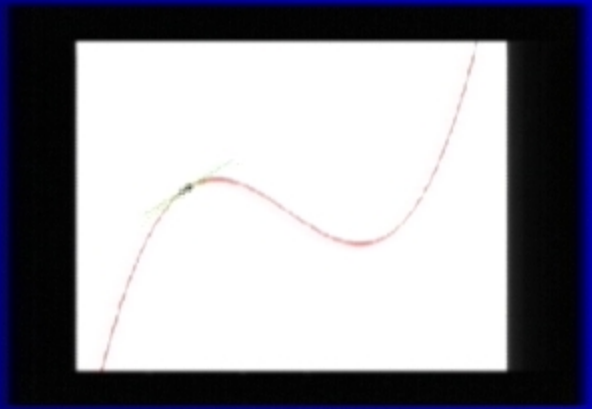


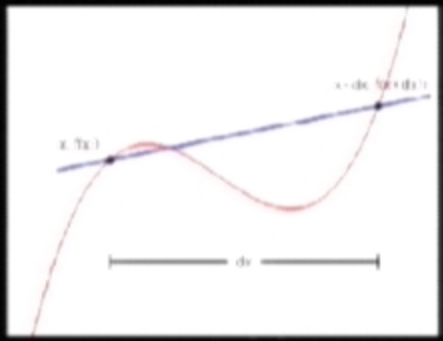

c^2
a^2
b^2


main
Func
Func
30

aligning same aspect






x + dx, f(x + dx)
x, f(x)
dx

Multimedia Annotation

David Bargeron
Interactive Visual Media Group
May 2002

Video Cliplets

Media Futures Team

Common Annotation Framework V1.0

- Platform for annotations
  - Schema and Object Model
  - Annotations are links
- Design Goals
  - Lightweight, extensible, and storage neutral
  - Universal Annotation Support
  - Facilitate Consistent User Experience
- Office.NET WordML schema integration
  - Universal Annotation Support
    - led by members of the Authoring Services Team (Chris Pratley, Brian Jones, Andrew Bishop,…)

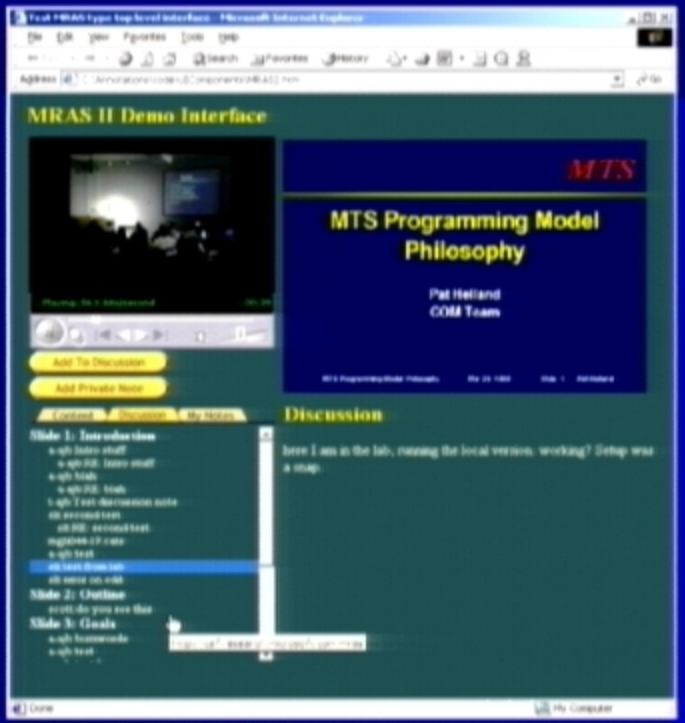
Personal/Shared Annotations

- In-context shared discussion
- Private/public annotations
- Robust anchors

- Robust anchors

Video Annotation

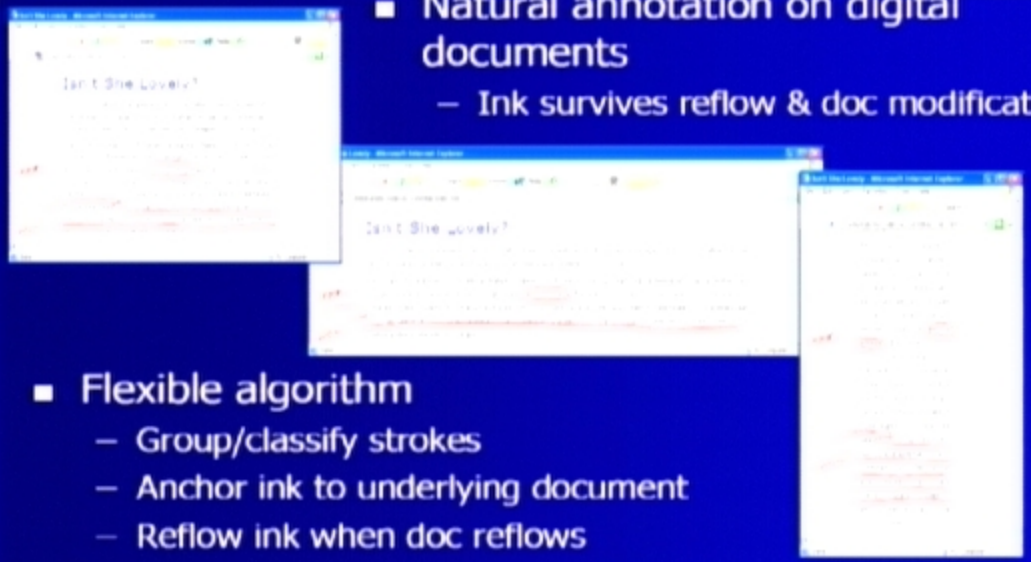
- Sharing and filtering
- Access Control
- TOC, discussion, personal notes, and slides are annotations



Common Annotation Framework V2.0

- Additional Goals
  - .NET compatibility
  - Strong schema types
  - UI Component Framework
- Designed with members of Active Reading team
  - Axel Kramer, Cathy Marshall, Ralph Ruiz
- Avalon integration, led by Avalon & eDocs Teams
  - Kevin Gjerstad, Ben Westbrook, Susi Strom, Dina Fesselmeyer
  - Out-of-the-box annotation functionality for Avalon apps
  - Avalon annotation platform available to other apps too

Digital Ink Annotations



- Natural annotation on digital documents
  - Ink survives reflow & doc modifications
- Flexible algorithm
  - Group/classify strokes
  - Anchor ink to underlying document
  - Reflow ink when doc reflows

Digital Ink Annotations

- Lots more to do…
  - Progressive learning to adapt to individual annotation style
  - "Actionable" annotations
  - Use ink annotations for search (with Zheng Chen's MarkIt!)
  - Capture ink annotations from paper (with Jian Wang/uPen)
  - Direct Ink Video Annotation (DIVA)
- Collaborators
  - Patrice Simard and Michael Shilman from MSR
  - RichInk team (Sashi Raghupathy, Matt Rhoten, Zoltan Szilagyi)
  - Tablet PC group (Bert Keely and many others)

Video Cliplets

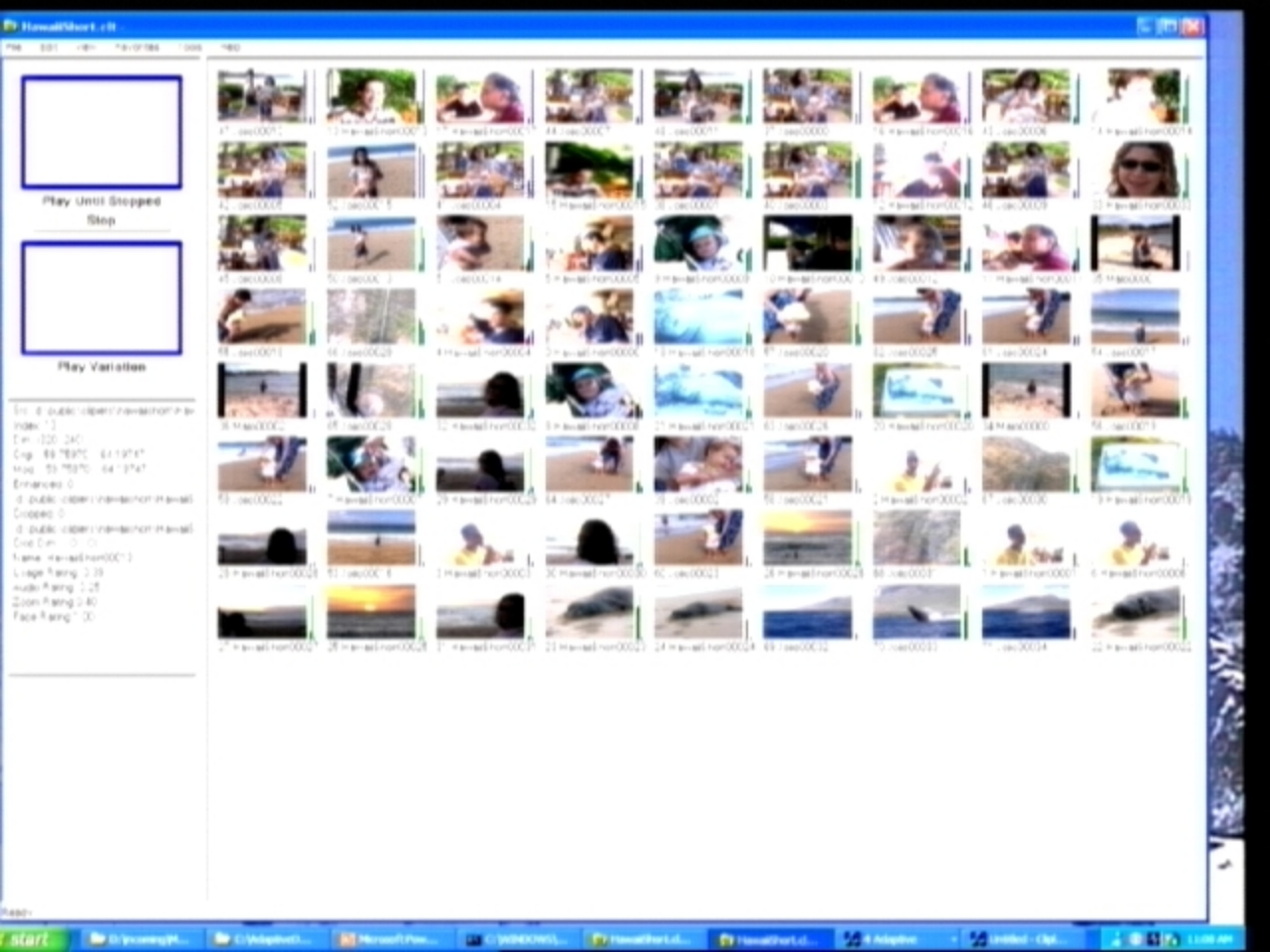
Media Futures Team

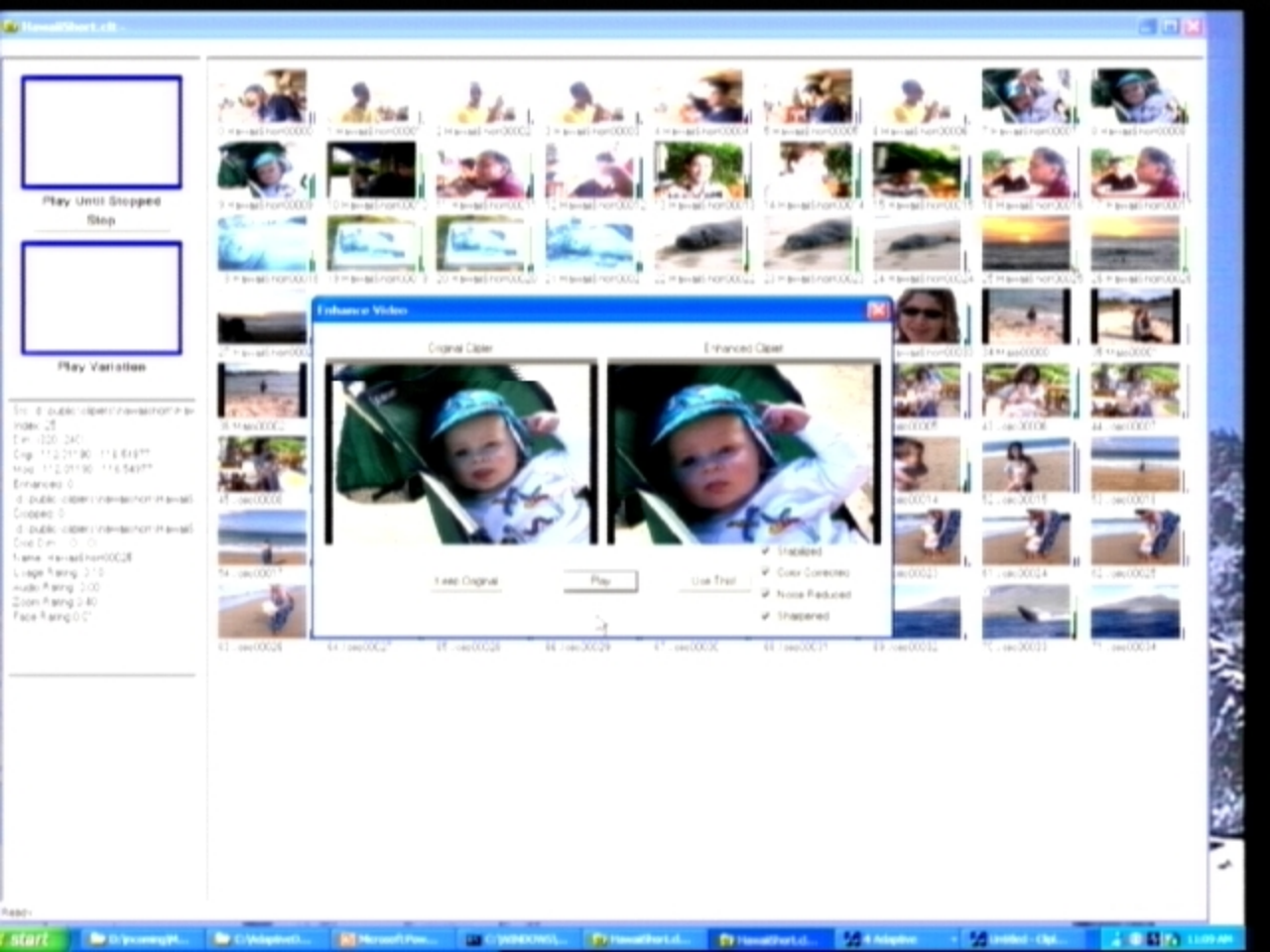
Project Contributors

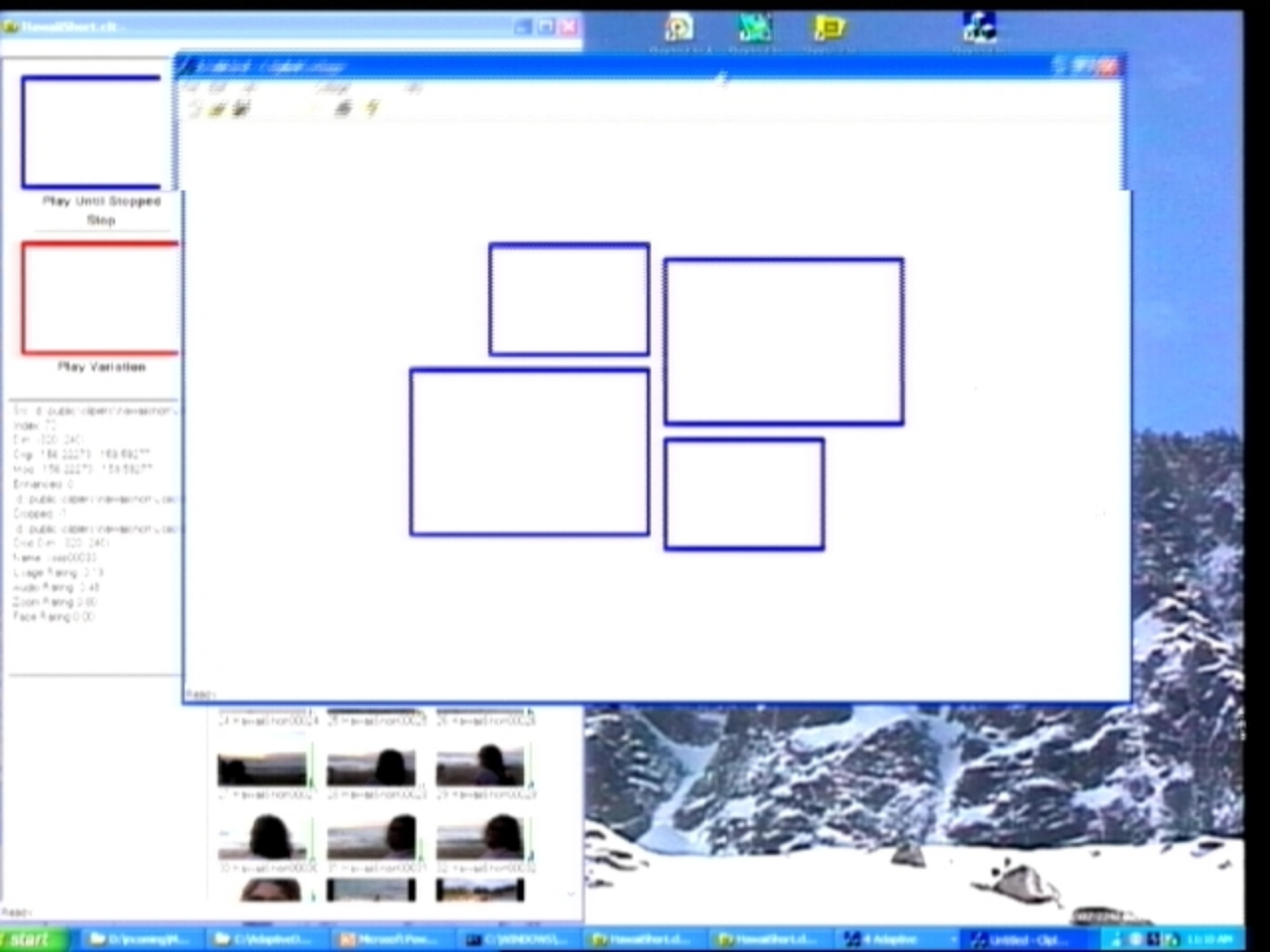
- Hagai Attias
- Dave Bargeron
- Steve Drucker
- Michel Gangnet
- Asta Glatzer
- Steve Harris
- Nebojsa Jojic
- Stan Li, *et al.*
- Patrick Perez
- Andreas Soupliotis
- Kentaro Toyama
- Jaco Vermaak

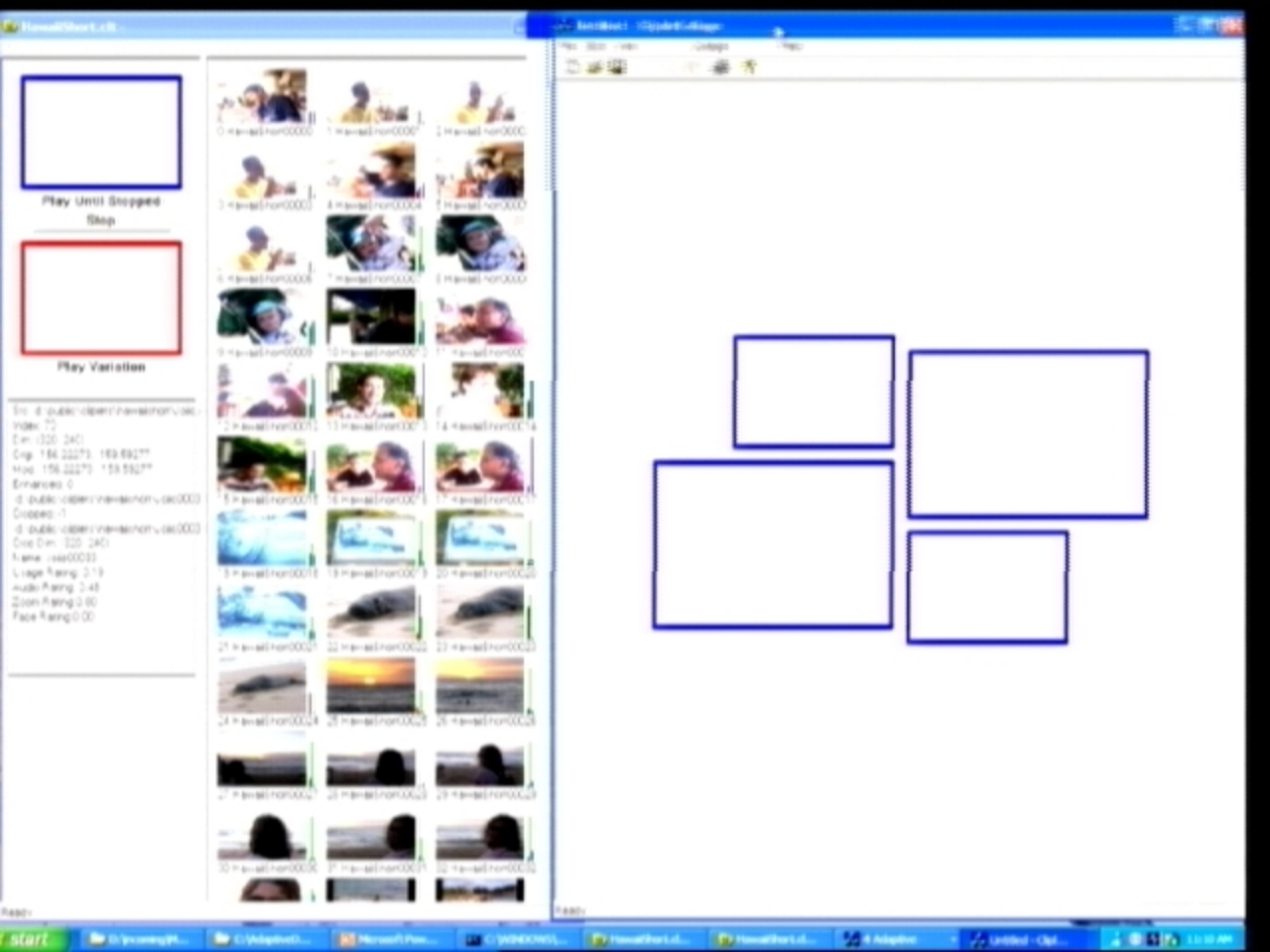
Video Cliplets

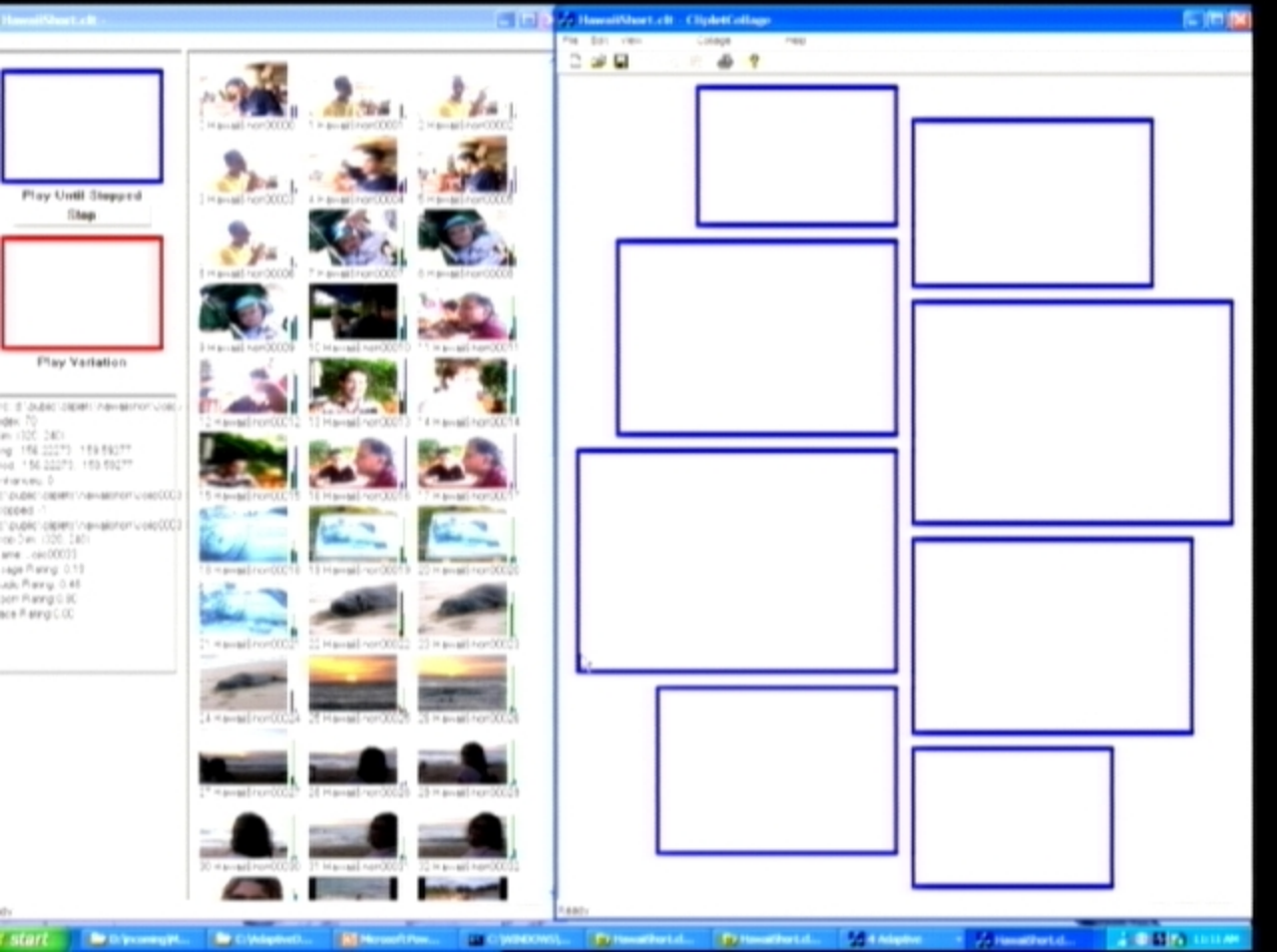
Media Futures Team

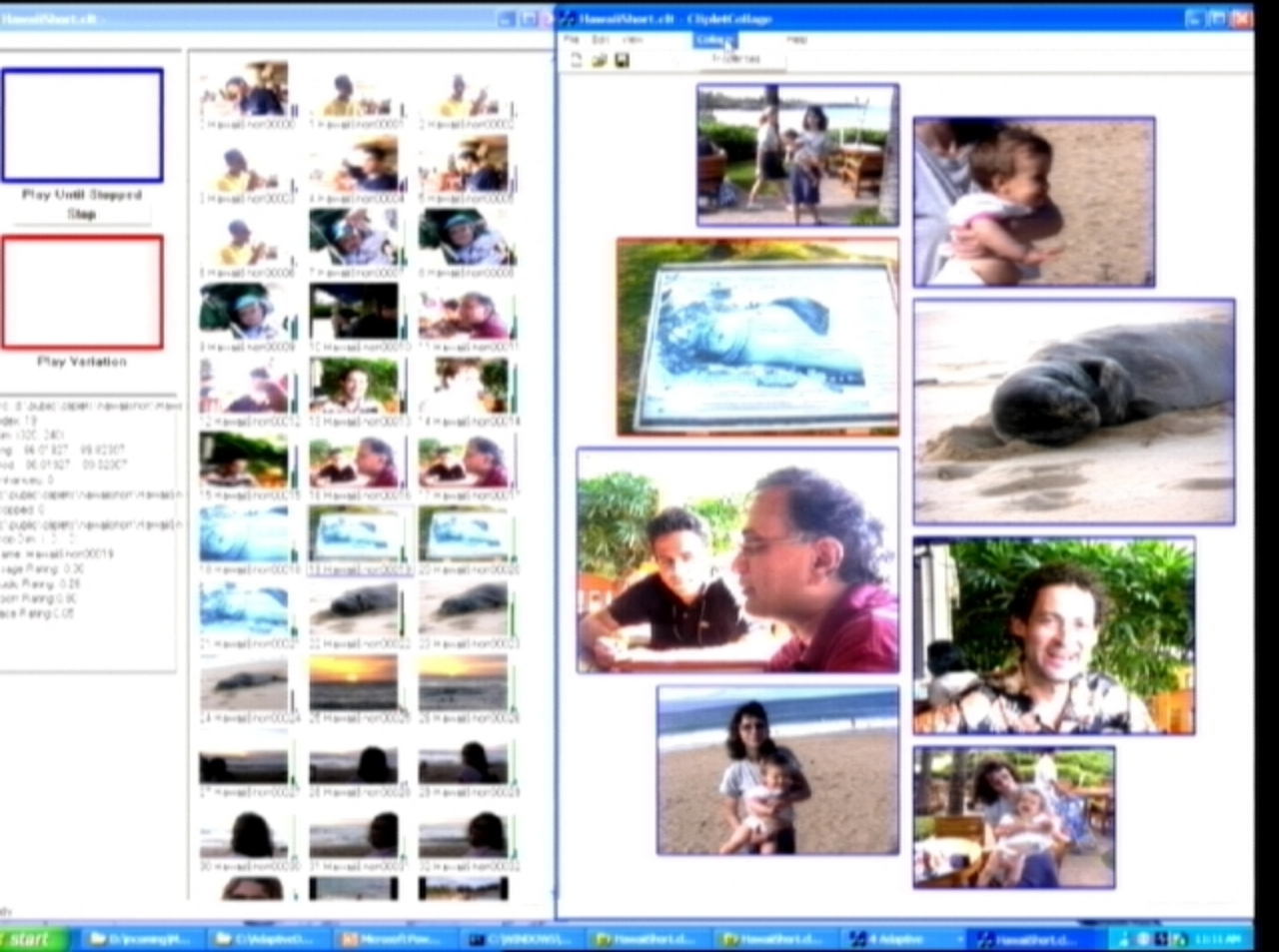

Play Until Stopped
Stop
Play Variation


Play Until Stopped
Stop
Play Variation
Enhance Video
Original Clip#
Enhanced Clip#
View Original
Play
Use This
Stabilized
Color Corrected
Noise Reduced
Sharpened


Play Until Stopped
Stop
Play Variation


Play Until Stopped
Stop
Play Variation


Play Until Stepped
Step
Play Variation


Play Until Stopped
Stop
Play Variation

Current/Future Work

- Cliplet beta application
  - User studies
- Representations of video
  - Internal
    - Metadata schemas
    - "Versioning" policies for video
  - Visible
    - UI for WinDB querying
    - UI for display of query results

Cliplets and Product Groups

- Working towards stronger video UX
  - Windows shell
  - Movie Maker / Producer
- Anticipating more interaction with…
  - WinFS (a.k.a. WinDB, Mighty Mouse)
  - Media Foundations
  - Digital Memory Project (now under WinFS)
  - WinFS (a.k.a. WinDB, Mighty Mouse)

Enhancing Consumer Video

Andreas Soupliotis

Video Enhancement

- Consumers are
  - Experts at watching produced video
  - Terrible at shooting own video

Video Enhancement

- Consumers are
  - Experts at watching produced video
  - Terrible at shooting own video
- Typical problems
  - Camera is shaking
  - CCD noise
  - Interlacing artifacts
  - Poor color balance
  - Out of focus

Stabilization

- Home video always contains handheld jitter
- Makes video look cheap and unprofessional
- The Blair Witch Effect!

Stabilization

- At the heart: image registration

Frame t

Frame t+1

Stabilization

- Image registration
  - Solve for transformation that aligns image t to t+1
  - Undo camera motion: warp image t+1 by the inverse transform

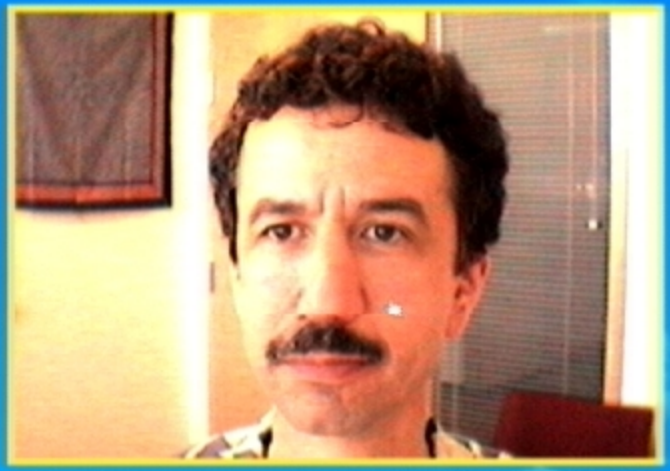
Frame t

Frame t+1 warped by inverse


PYRO
videoWAVE

Denoising

- Digital video cameras introduce "CCD noise"
- Goal: remove noise, but keep video sharp

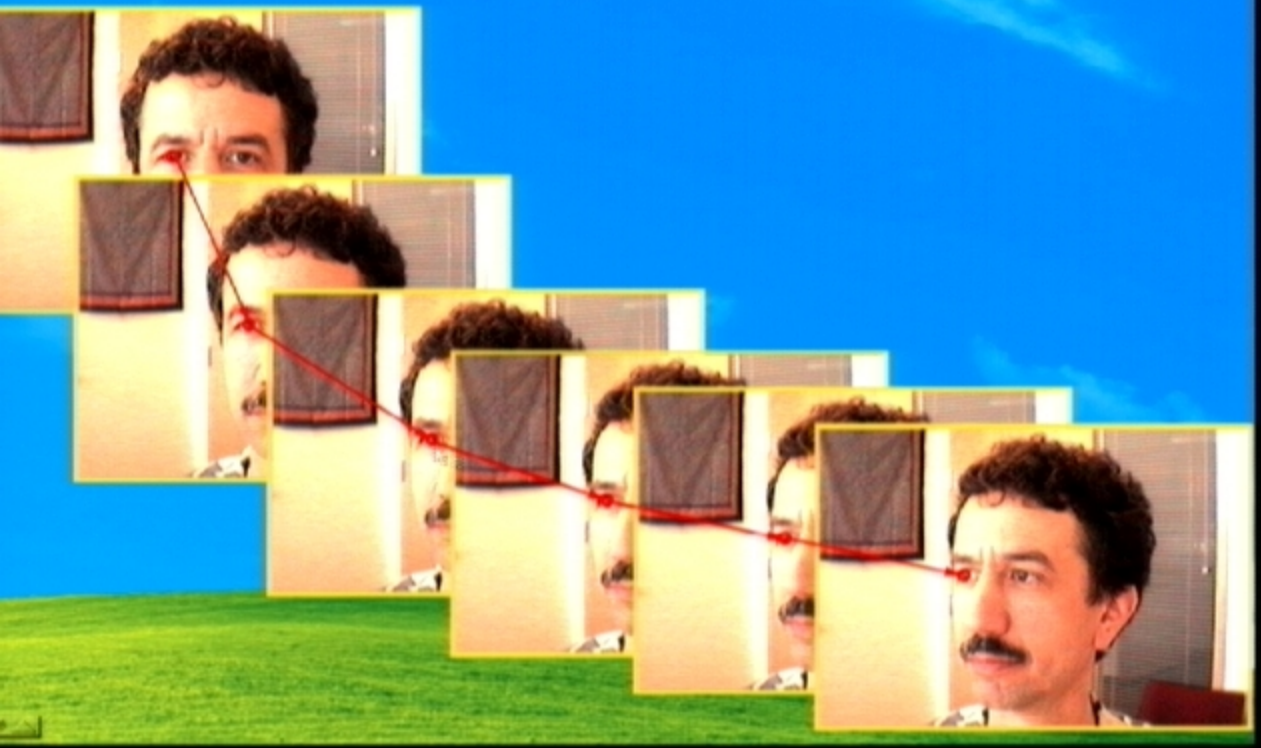


Denoising

- Temporal filtering

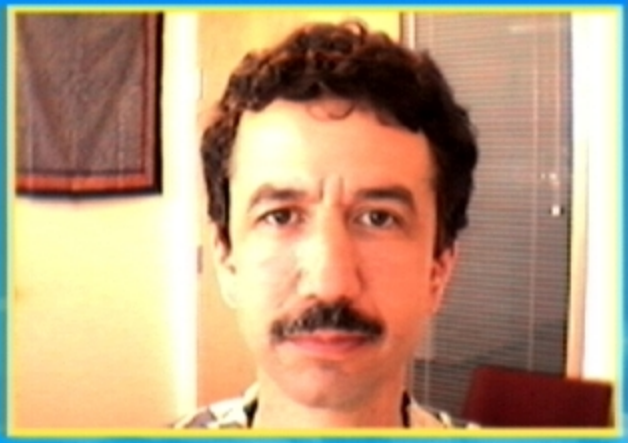
Denoising

Denoising: Optic Flow



Denoising

- The result



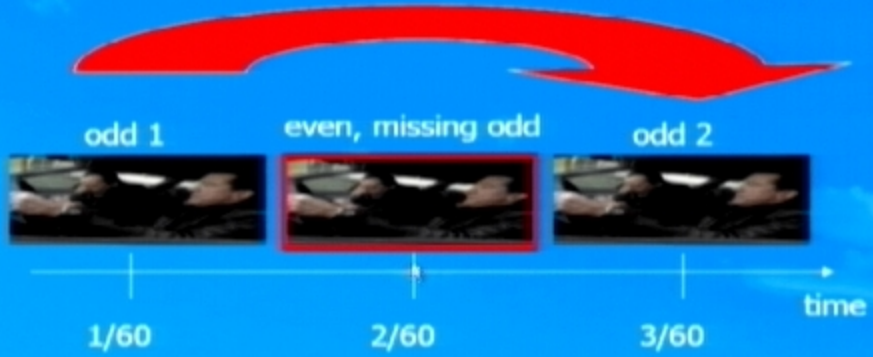
Slow Motion

Slow Motion with Optic Flow

Extracting a Still From Interlaced Video

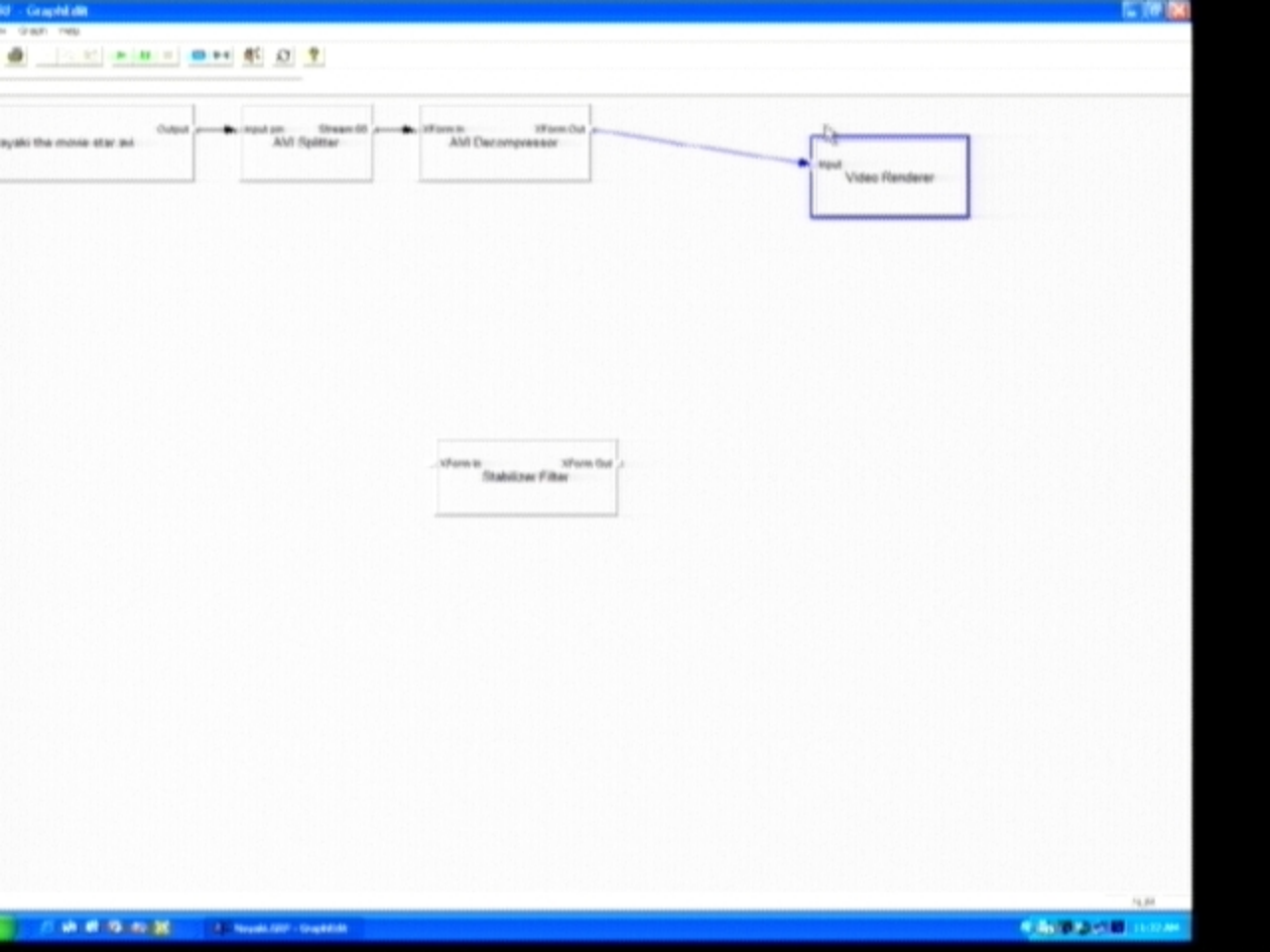
Extracting a Still from Interlaced Video

- Optic flow to the rescue again



Technology Transfer Path

- Real-time stabilization: MovieMaker & Longhorn Shell
- Real-time optic flow
- Motion information as metadata
- Slow motion
- High resolution still extraction


Output
Input pin
Stream 00
AVI Splitter
XForm In
XForm Out
AVI Decompressor
Input
Video Renderer
XForm In
XForm Out
Stabilizer Filter

Enhancing Consumer Video

Andreas Soupliotis

Collaborators

- Brendan Frey
- Anitha Kannan
- Nemanja Petrovic
- Hagai Attias
- Matt Beal

Collaborators

- Brendan Frey
- Anitha Kannan
- Nemanja Petrovic
- Hagai Attias
- Matt Beal

Overall goal

- Analyze video sequences to find
  - Objects and sounds in video
  - The rules governing their appearance changes
  - Motion and sound patterns
  - Higher level properties of their behavior
  - Audio-visual correlations, etc.
- Applications:

  summarization, enhancement, retrieval, watermarking, compression,…

Inverting image formation

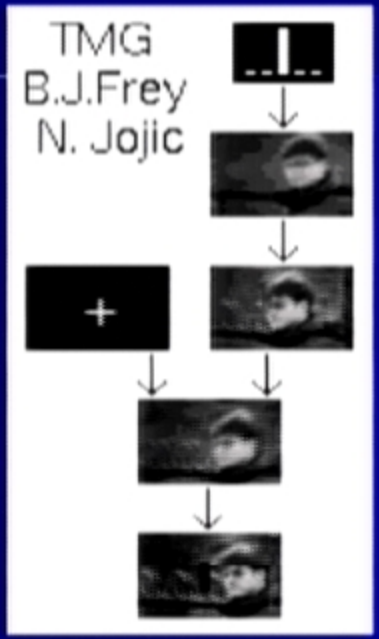
- Generative density model
  - reflects desired structure
  - *randomly generates* plausible images,
  - represents the data by *parameters*
- Inverting it is possible!

(Ultimately, one must treat all causes of variability jointly, in an unsupervised manner in order to solve vision)

Scene- and object-based models

- Transformed mixtures of Gaussians
  - Mean and variance map for an object or a scene
  - Transformation modeling the scene motion
  - Clustering frames of video by visual similarity
- Scene+blobs model
  - One scene model as above
  - Several moving blobs capturing the rest of the object in the scene
  - Transformation modeling the camera motion
  - Clustering or example-based search (Intelligent Fast-Forward)
- Layered flexible sprites
  - Multiple objects with means and variances in both appearance and shape
  - Transformations in each layer modeling each object's motion
  - Video editing

TMG: Fitting a generative model



Class index

Class mean (representative image)

Mean with added variability

Transformed (shifted image)

Transformed image with added non-uniform noise

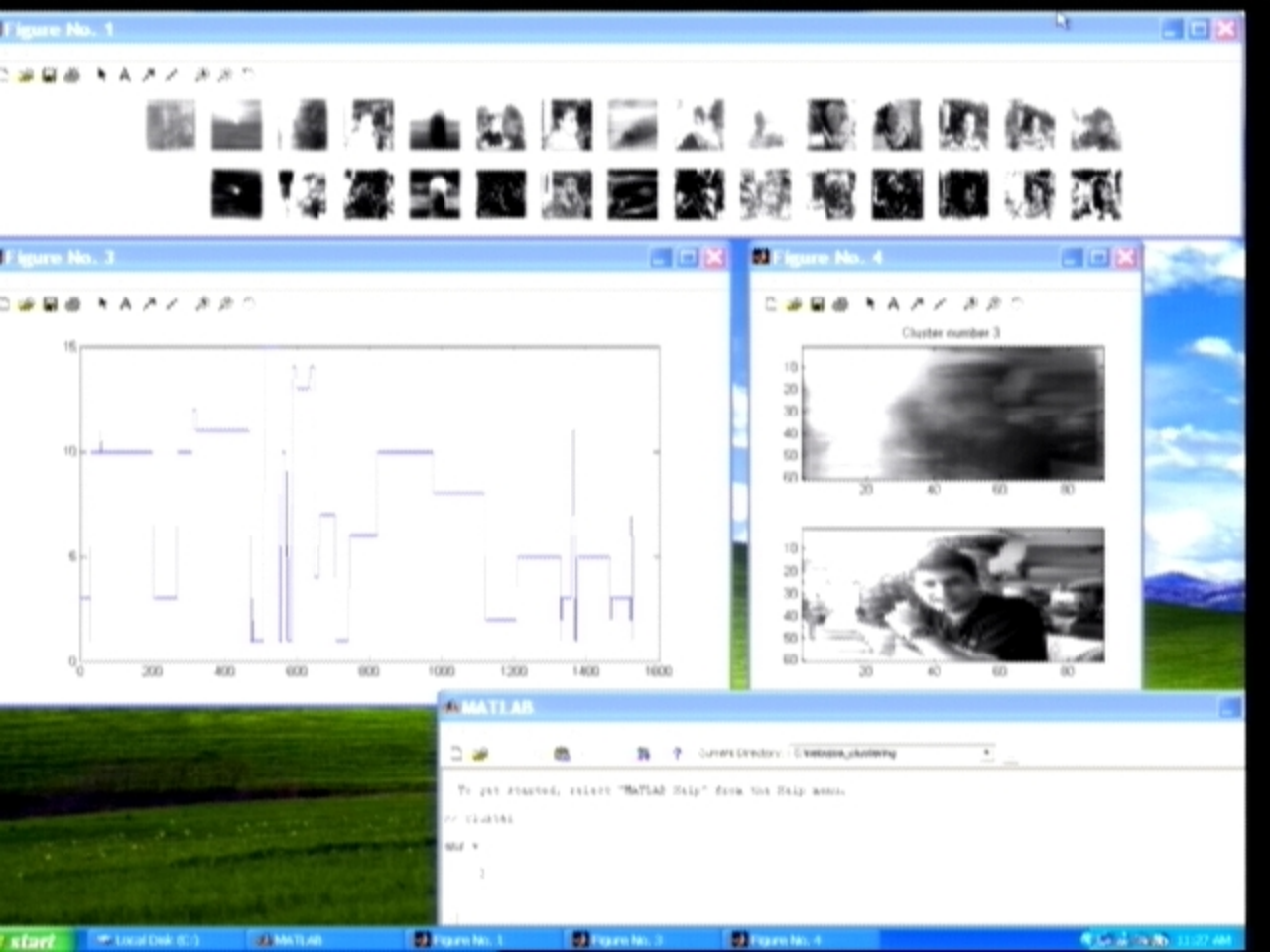
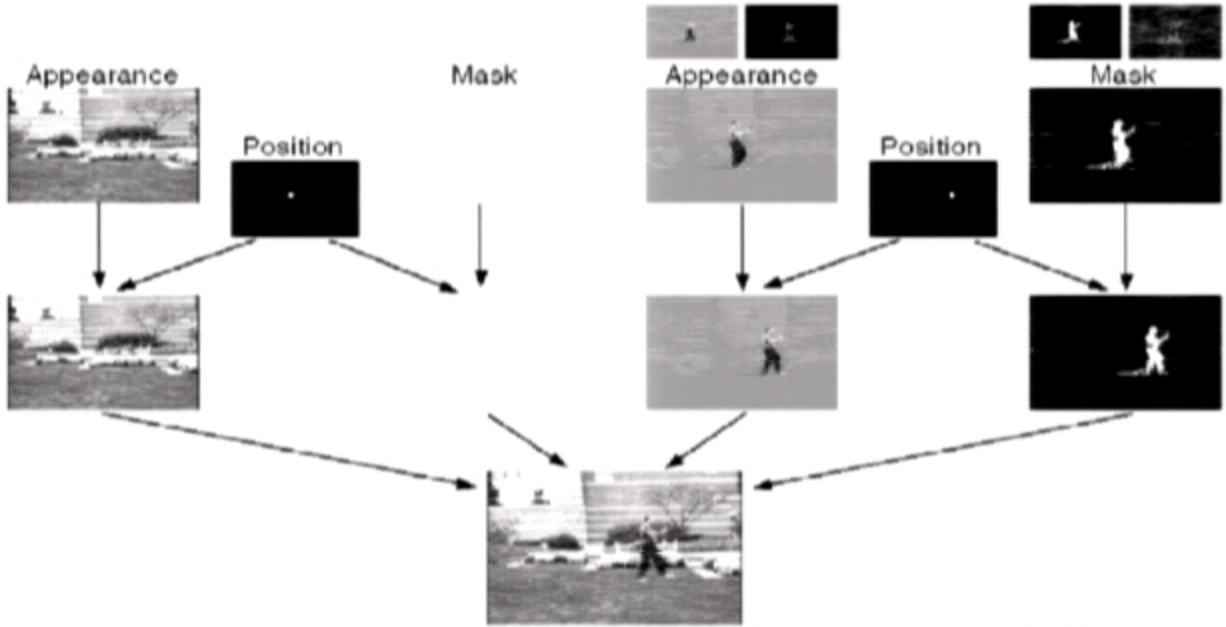


Figure No. 1
Figure No. 3
Figure No. 4
Cluster number 3
MATLAB
start

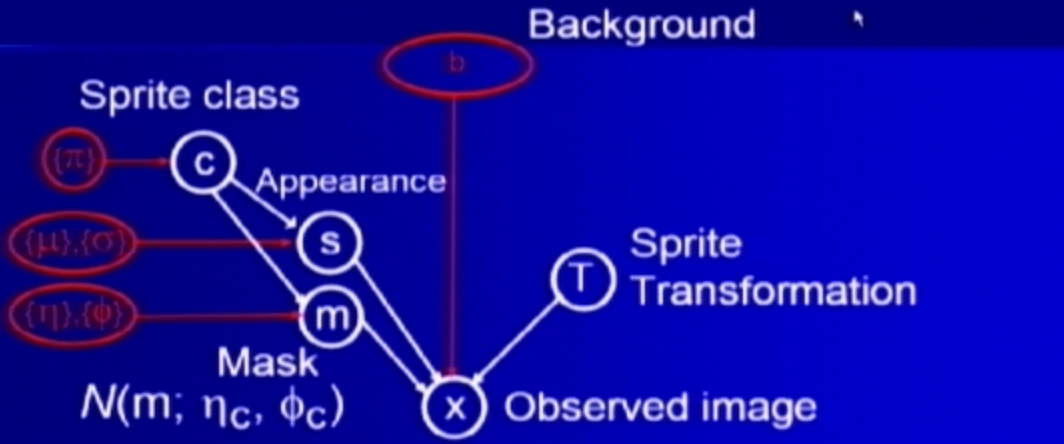
Moon-walking

Layered sprites, Jojic & Frey 2001



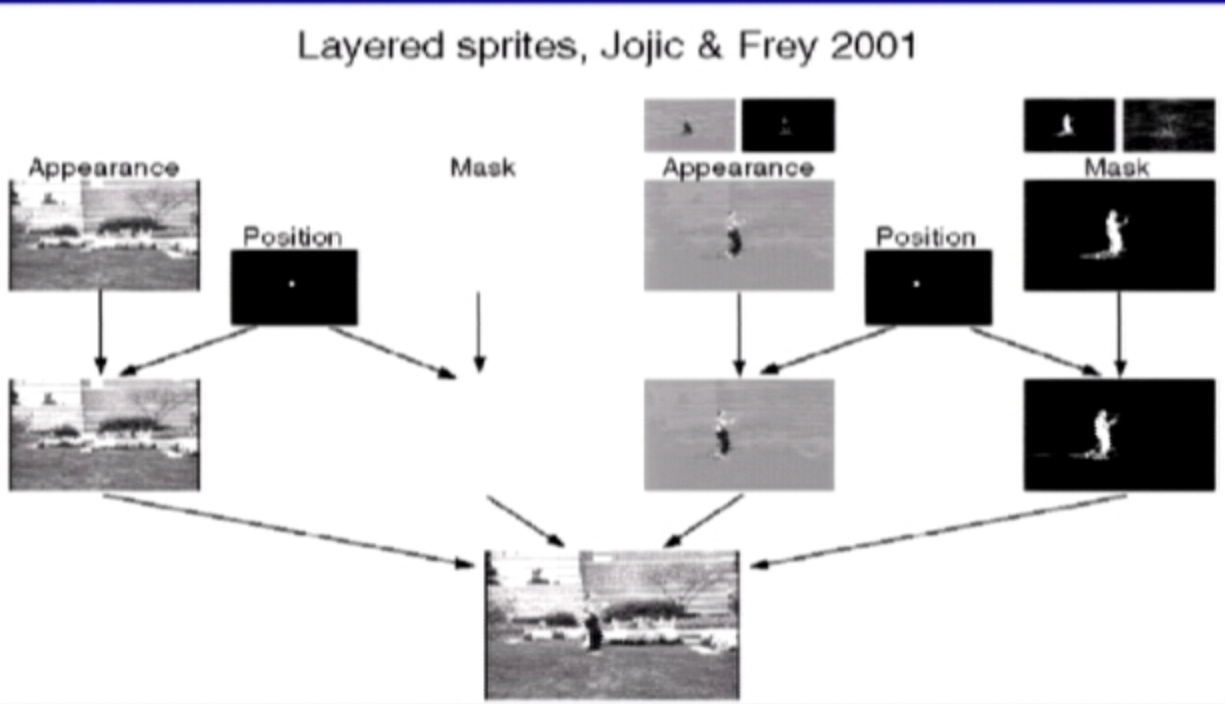
Parsing Video into Layers

Layered sprite model



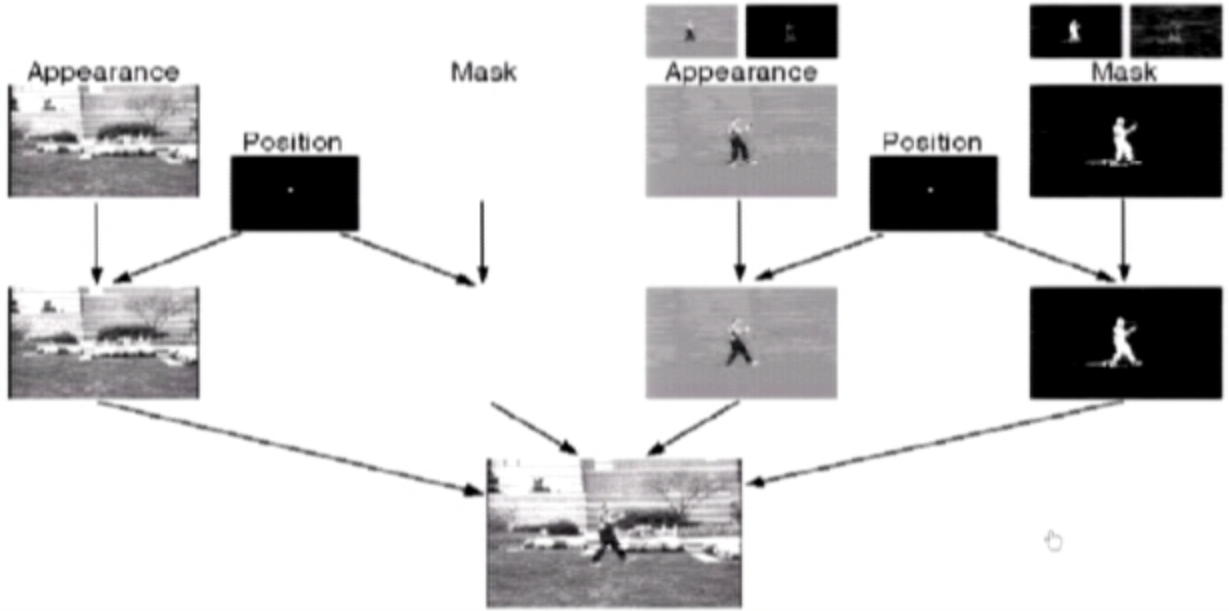
Input video

Basic flexible layer model



Stabilization

Layered sprites, Jojic & Frey 2001



Object removal

What Happens If There Aren't Layers

From the New Yorker

What Happens If There Aren't Layers

From the New Yorker

Comparisons

- Transformed mixtures of Gaussians:
  - 3 frames per second (for clustering)
  - Can do scene clustering or object clustering if objects are large
  - No initialization, no pre-set parameters
- Scene+object model
  - 10 frames per second (for training)
  - 120 frames per second (search)
  - Robust to occlusions, can distinguish between scenes with the same background but different foreground
  - No initialization, no pre-set parameters
- Layered flexible sprites
  - Seconds per frame
  - Can subtract objects, compute panoramas of dynamic scenes
  - No initialization, no pre-set parameters

Work in progress: Applications

- Fast DShow filter for frame clustering
- Clustering faces in photos
- Click passwords
- FaceCerts – secure photo IDs

Image Based Realities

Rick Szeliski, Sing Bing Kang, Matt Uyttendaele, Antonio Criminisi

Virtual Home Tour

Image-Based Realities Team

P. Anandan, Antonio Criminisi, Sing Bing Kang,
Rick Szeliski, Matt Uyttendaele

http://www.research.microsoft.com/vision/ImageBasedRealities/

Project goals

Project goals

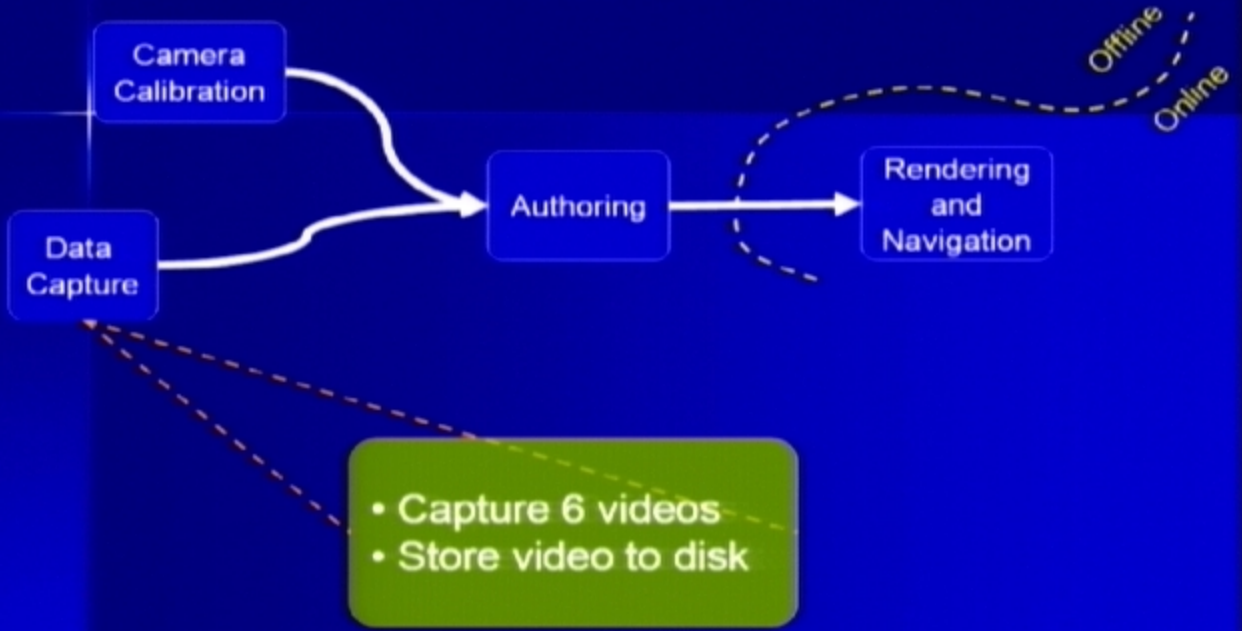
Goal: *Combine the **interactivity** of 3D games and virtual worlds with the **realism**, richness, and narrative of film and video*

Project goals

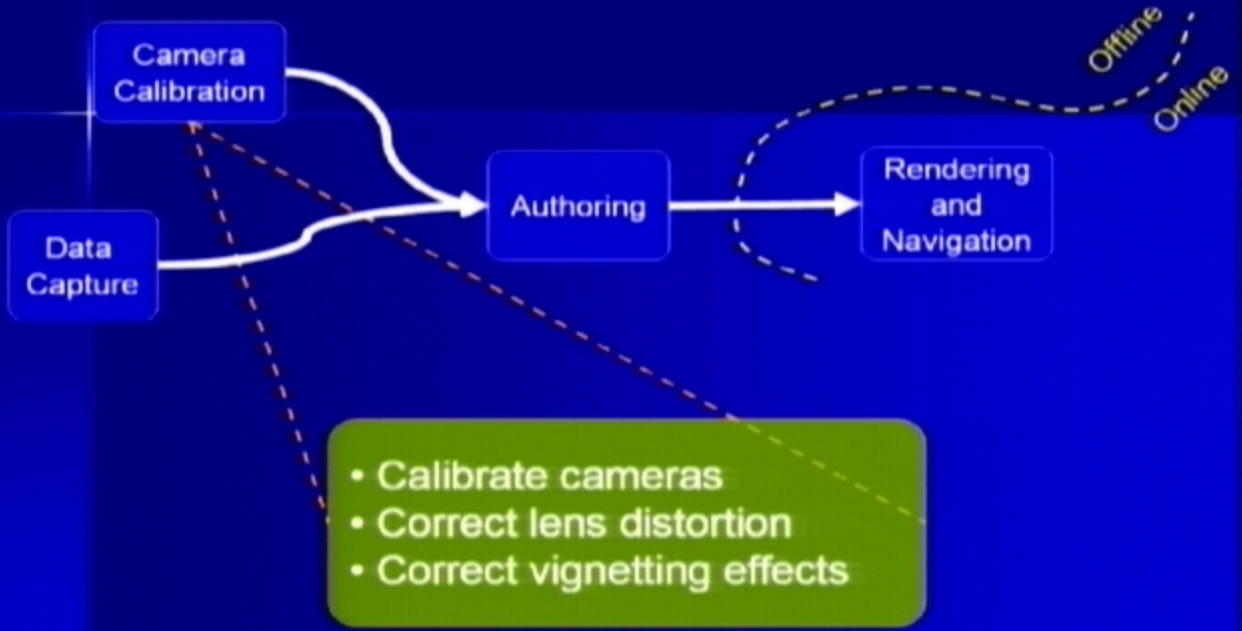
Goal: *Combine the **interactivity** of 3D games and virtual worlds with the **realism**, richness, and narrative of film and video*

Demo: *Capture real-world video footage with omni-directional camera and add interactive video/graphic elements*

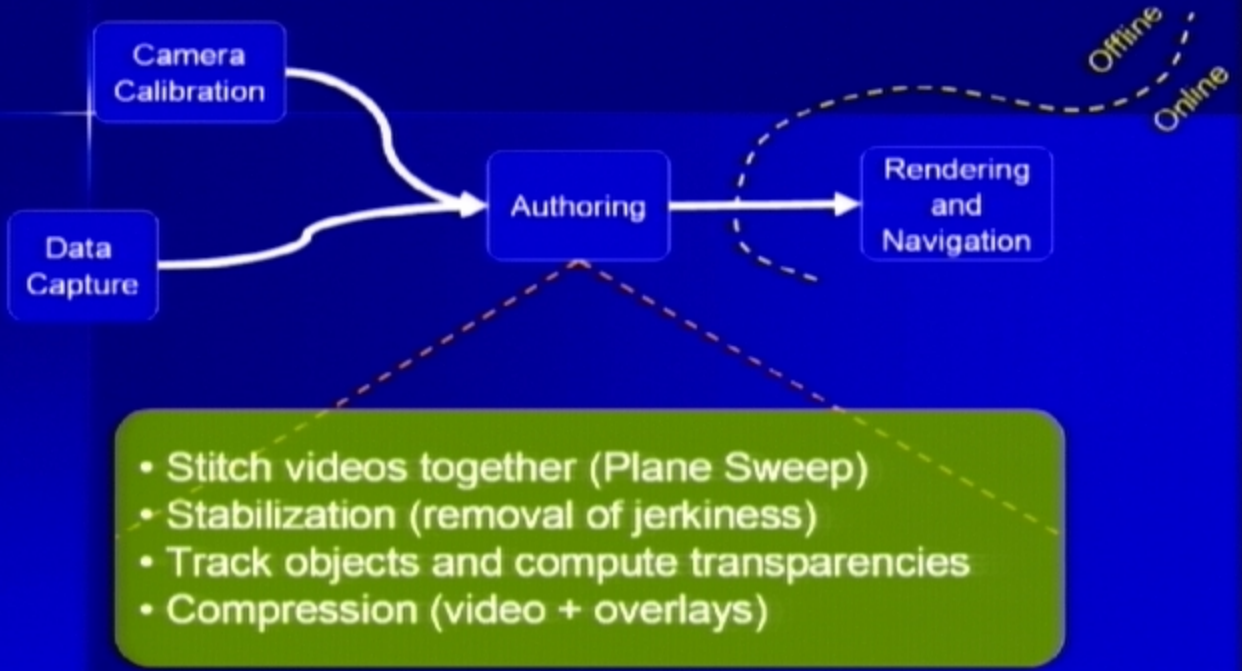
Overview


Camera Calibration
Data Capture
Authoring
Rendering and Navigation
Offline
Online
• Capture 6 videos
• Store video to disk

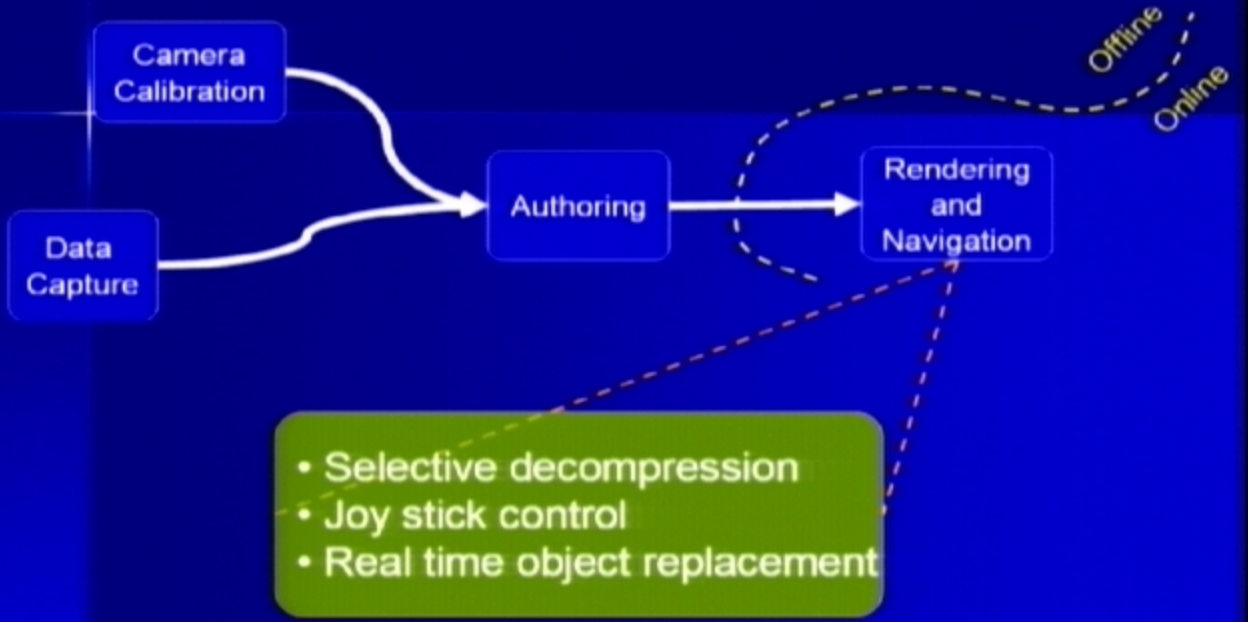
Overview


Camera Calibration
Data Capture
Authoring
Rendering and Navigation
Offline
Online
• Calibrate cameras
• Correct lens distortion
• Correct vignetting effects

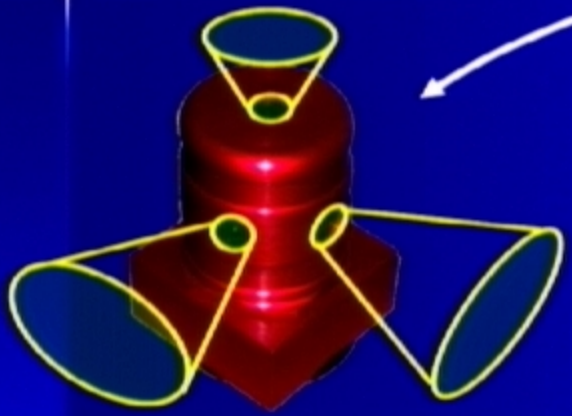
Overview


Camera Calibration
Data Capture
Authoring
Rendering and Navigation
Offline
Online
• Stitch videos together (Plane Sweep)
• Stabilization (removal of jerkiness)
• Track objects and compute transparencies
• Compression (video + overlays)

Overview

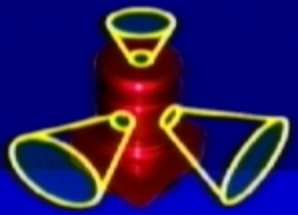

Camera Calibration
Data Capture
Authoring
Rendering and Navigation
Offline
Online
• Selective decompression
• Joy stick control
• Real time object replacement

Surround video acquisition system



OmniCam (six-camera head)

OmniCam

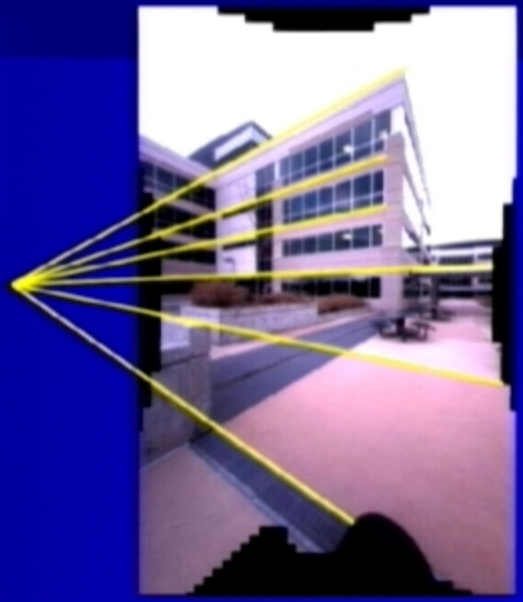


- Built by Point Grey Research
- Six camera head (*Ladybug*)
- Portable hard drives, fiber-optic link
- Resolution per image: 1024 x 768
- FOV: ~100° x ~80°
- Acquisition: 20 fps uncompressed

Original (distorted) image

Output (corrected, perspective) image

Calibration: Lens distortion



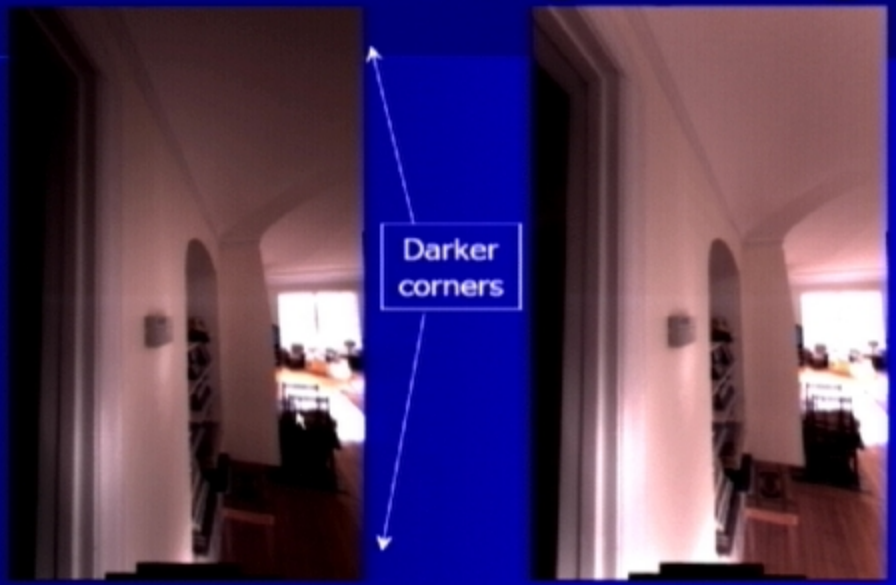
Original (distorted) image

Output (corrected, perspective) image

Calibration: Camera Intrinsics & Extrinsics

- Currently a simple model:
  - Focal Length / camera
  - 3D Rotation / camera
- Capture panorama of scene at "infinity" (ignore translations)
- Stitch scene in VideoMosaic
- Capture panorama of scene at

Calibration: Vignetting



Before correction

After correction

Calibration: Vignetting

Find f, α

Fit radial function

$$I(r) = I_0(1+(r/f)^2)^{-2}\,(1-\alpha r)$$

Optical vignetting

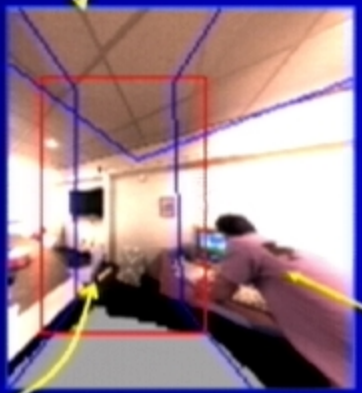
Geometric vignetting

Before vignetting correction

After vignetting correction

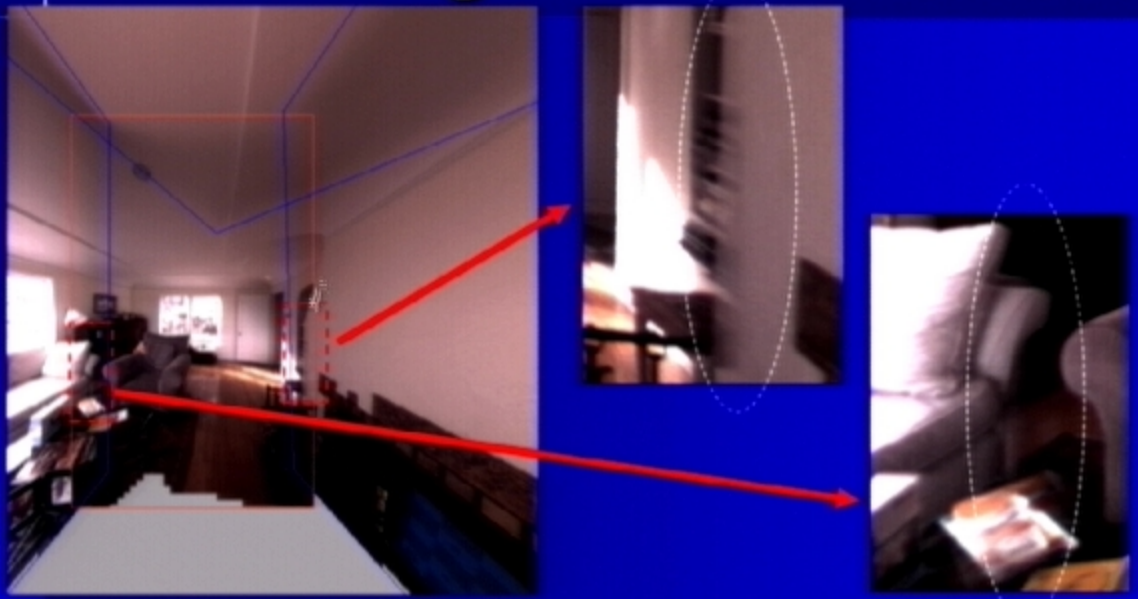
Authoring Software

Stitching



(Only 4 of 6 images shown here)

Problem with simple feathering

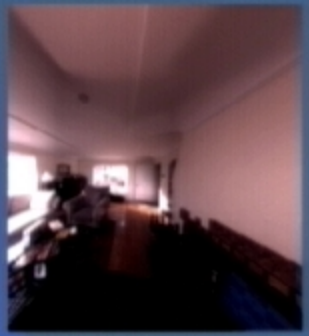


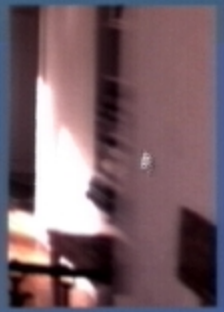
Our solution

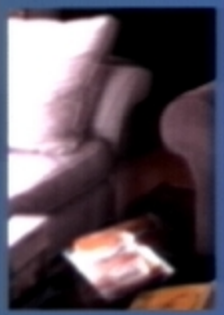
- Multiperspective plane sweep (MPPS)
- Apply to pairwise overlapped regions
- Steps:
  - Rectify
  - Plane sweep (multiple virtual cameras)
  - Blend colors

Stitching example

Without MPPS







With MPPS

Stabilize

Align frame-to-frame and distribute Δheading

Before motion stabilization

After motion stabilization

Augmented surround video

A movie clip from the original surround data (with occlusion).

The augmented movie clip. Occlusions are handled correctly.

Augmented surround video

A movie clip from the original surround data.

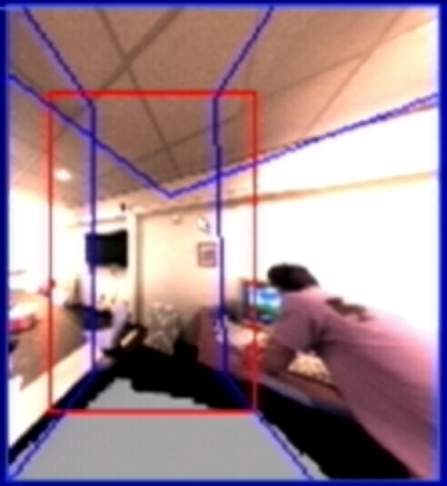
The augmented movie clip. Surround augmented with video

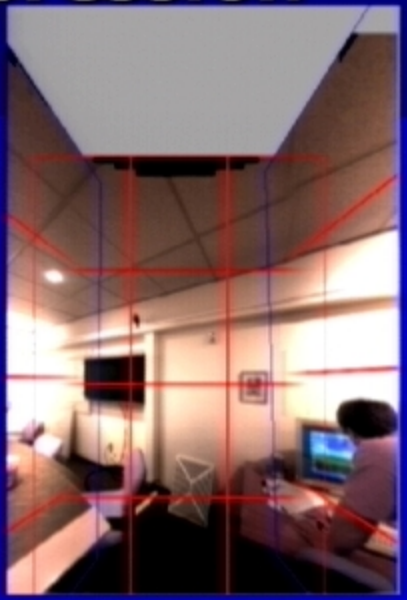
Viewing Software

Media Viewer

- Research Media Viewer:
  - D3D + DirectShow + XML scene graphs
- Model: 6 texture mapped planes
  - Rendered using D3D hardware
- New components:
  - Selective decompression
  - Joystick controller
  - Overlays support

Selective decompression





Observation: at most 4 inputs visible at one time (for reasonable fields of view)

1. Compress the tiles
2. Decompress and transfer (to texture memory) only visible tiles

Future work

- Camera Calibration

Future work (con't)

- Compression
  - 3D reconstruction
  - Lightfield compression
- Rendering
  - .Net architecture (Avalon 3D)
  - Directional sound

High dynamic range

- Combine multiple exposures

Underexposed

- Challenging for moving images (video)

High dynamic range

- Combine multiple exposures

Overexposed

- Challenging for moving images (video)

High dynamic range

- Combine multiple exposures

Combined

- Challenging for moving images (video)

Automatic filling of bottom view

Original (unfilled)

Filled with data borrowed from other frames

Exploit the huge redundancy of data to fill in the unseen portions

(Preliminary results)

Geometric reconstruction

Points tracked *in many frames in many views*

Edges tracked *in many frames in many views*

This is the first step towards complete 3D reconstruction …

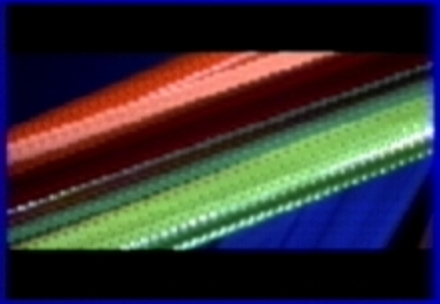
(Preliminary results)

Lightfield compression

- Generalization/extension of video compression
  - decompress in any direction
  - selective decompression
  - static world (depth compensation)
  - novel viewpoint generation
  - specularities/reflections

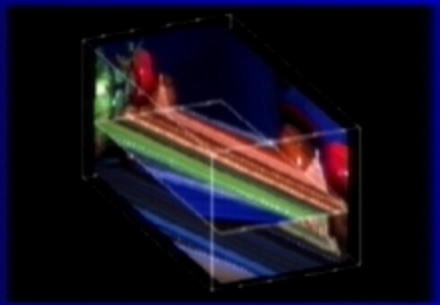
Lightfields and EPIs

video



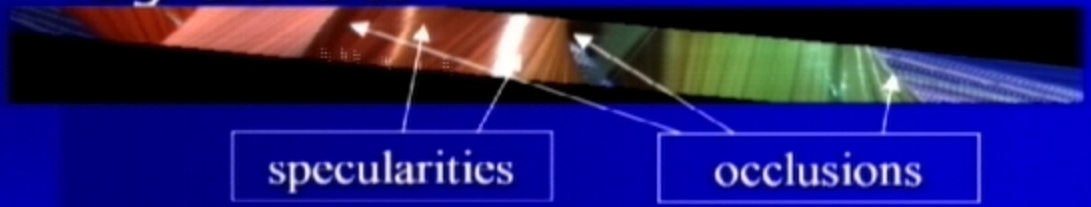
EPI

ST cube:



Sheared EPI

- Coherent trails of colors (light): *LightBundles*



Correspond to coherent surfaces/reflections

- Specular trails are (usually) lightly curved

[R. Swaminathan *et al.*, ECCV'2002]

Rendering architecture

- Avalon 3D/Video
  - Extend Avalon 2D rendering architecture to 3D and video
    - XML + SMIL + D3D + DShow
  - Joint work with Graphics (Charles Loop et al.)
  - New substrate for IBR research

Novel acquition platforms

- Robotic cart

Image-Based Realities

- Contributions
  - Novel sensors for "ultimate reality"

Image-Based Realities

- Contributions
  - Novel sensors for "ultimate reality"
  - Analysis techniques for interactions with the "virtualized world"

Image-Based Realities

- Contributions
  - Novel sensors for "ultimate reality"
  - Analysis techniques for interactions with the "virtualized world"
  - Novel lightfield compression
  - Advanced image-based rendering software and architecture
  - Multimedia merger: games + video

Thank you!

The IBR team

Three factors

- The growth of the imaging and graphics capabilities on the PC
- The explosion of consumer end digital image and video acquisition
- The increase in connectivity

Consequences

- A visually rich media experience on the PC
- Breakdown of the barriers for seamless media integration
- A fundamental paradigm shift toward interactive media

How to Get There?

- Technologies for media analysis and synthesis
- Novel, visually rich, and interactive user experiences
- Changes to the Platform
  - Interactive media (especially video)
  - Seamless media integration

Strategic Initiatives

- Seamless Multimedia Integration
  - (with Graphics and Nextmedia groups)
  - Avalon 3D, Shell MSX, Sparkle, DMD
- Interactive Video
  - (with Graphics, Nextmedia, Signal Proc)
  - Media Foundation, Audio-video Platform Group
- First class personal video
  - Windows Media Tools, Shell MSX

Strategic Initiatives

- Seamless Multimedia Integration
  - (with Graphics and Nextmedia groups)
  - Avalon 3D, Shell MSX, Sparkle, DMD
- Interactive Video
  - (with Graphics, Nextmedia, Signal Proc)
  - Media Foundation, Audio-video Platform Group
- First class personal video
  - Windows Media Tools, Shell MSX

Interactive Visual Media Group is:

P. Anandan
Antonio Criminisi
Nebojsa Jojic
David Salesin
Andreas Soupliotis
Matt Uyttendaele

David M. Bargeron
Chuck Jacobs
Sing Bing Kang
Rick Szeliski
Kentaro Toyama